Bayesian calibration at the urban scale

A case study on a large residential heating demand application in Amsterdam

Wang, C.; Tindemans, Simon; Miller, Clayton; Agugiaro, Giorgio; Stoter, Jantien

DOI
10.1080/19401493.2020.1729862

Publication date
2020

Document Version
Final published version

Published in
Journal of Building Performance Simulation

Taylor & Francis
Taylor & Francis Group

Journal of Building Performance Simulation

ISSN: 1940-1493 (Print) 1940-1507 (Online) Journal homepage: https://www.tandfonline.com/loi/tbps20

Bayesian calibration at the urban scale: a case study on a large residential heating demand application in Amsterdam

Cheng-Kai Wang, Simon Tindemans, Clayton Miller, Giorgio Agugiaro & Jantien Stoter

To cite this article: Cheng-Kai Wang, Simon Tindemans, Clayton Miller, Giorgio Agugiaro & Jantien Stoter (2020) Bayesian calibration at the urban scale: a case study on a large residential heating demand application in Amsterdam, Journal of Building Performance Simulation, 13:3, 347-361, DOI: 10.1080/19401493.2020.1729862

To link to this article: https://doi.org/10.1080/19401493.2020.1729862

JOURNAL OF BUILDING PERFORMANCE SIMULATION
2020, VOL. 13, NO. 3, 347–361
https://doi.org/10.1080/19401493.2020.1729862

Bayesian calibration at the urban scale: a case study on a large residential heating demand application in Amsterdam

Cheng-Kai Wang [a], Simon Tindemans [b], Clayton Miller [c], Giorgio Agugiaro [a] and Jantien Stoter [a]

[a]3D Geoinformation Group, Delft University of Technology, Delft, Netherlands; [b]Department of Electrical Sustainable Energy, Delft University of Technology, Delft, Netherlands; [c]Building and Urban Data Science Group, National University of Singapore, Singapore, Singapore

ABSTRACT

A bottom-up building energy modelling at the urban scale based on Geographic Information System and semantic 3D city models can provide quantitative insights to tackle critical urban energy challenges. Nevertheless, incomplete information is a common obstacle to produce reliable modelling results. The residential building heating demand simulation performance gap caused by input uncertainties is discussed in this study. We present a data-driven urban scale energy modelling framework from open-source data harmonization, sensitivity analysis, heating demand simulation at the postcode level to Bayesian calibration with six years of training data and two years of validation data. Comparing the baseline and the calibrated simulation results, the averaged absolute percentage errors of energy use intensity in the study area have significantly improved from 25.0% to 8.3% and from 19.9% to 7.7% in two validation years, while $CVRMSE_{2016} = 11.5\%$ and $CVRMSE_{2017} = 13.2\%$. The overall methodology is extendable to other urban contexts.

ARTICLE HISTORY
Received 7 July 2019
Accepted 10 February 2020

KEYWORDS
Urban building energy modelling; simulation performance gap; geographic information system; sensitivity analysis; Bayesian calibration; spatial-temporal modelling

1. Introduction

The building sector plays an indispensable role in achieving a low-carbon future as it accounts for more than one-third of total final energy use and greenhouse gas emissions (International Energy Agency 2013). Meeting increased energy demand while achieving decarbonization targets in rapidly urbanizing societies has become a common challenge faced by most cities around the world. Frayssinet et al. (2018) breaks down urban energy challenges into three aspects, which illustrate why urban energy modelling has remained an active field for the last 30 years (Keirstead, Jennings, and Sivakumar 2012).

(1) The urban population is rapidly increasing, with 54% of the population living in cities in 2014, which is expected to climb to 66% by 2050 (United Nations 2014). This fact is accompanied by increased energy demand per capita by 32% in the last 40 years (International Energy Agency 2013). Sustainable urban development and holistic energy policies to balance the increased resource demands are thus crucial issues.
(2) Renewable energy sources are changing the landscape of energy market rapidly. Because of the decentralized and intermittent characteristics of renewable energy sources, a comprehensive understanding of urban energy systems is crucial to bridge the gap between energy demand and supply (International Energy Agency 2014).
(3) Urban heat stress and increased cooling demands caused by Urban Heat Island (UHI) effects may be further intensified and become more frequent in the context of climate change (Intergovernmental Panel on Climate Change 2007) and lead to increasing public health problems.

To address these challenges ahead, the building sector has tremendous potential to reduce energy consumption and greenhouse gas emissions by improving building energy efficiency such as enforcing appropriate retrofit measures and adopting energy-neutral buildings design and urban districts planning (International Energy Agency 2013). In this prospect, urban scale building energy modelling (UBEM) can be a crucial enabler to depict the current state of urban energy demand and predict its future evolution (Reinhart and Davila 2016).

Two urban building energy modelling approaches can be generally distinguished: top-down and bottom-up methods, according to Swan and Ugursal (2009). The top-down approach models long-term total energy demand of the building stock based on macroeconomic and socioeconomic parameters. It has a relatively coarse spatial-temporal resolution, often at the national scale, of energy consumption of the building stock. On the other hand, the bottom-up approach is more capable of energy modelling at disaggregated scales.

The bottom-up approach can be further separated into a statistical method and engineering method (Swan and Ugursal 2009; Kavgic et al. 2010). The bottom-up analytical approach (Guerra Santin, Itard, and Visscher 2009; Howard et al. 2012; Mastrucci et al. 2014; Nouvel et al. 2015; Torabi Moghadam et al. 2018) relies on historical consumption data, and building stock characteristics data, which are often derived from Geographical Information System (GIS) and can be further enriched

CONTACT Cheng-Kai Wang ckwang25@gmail.com

with local (census) statistics, to build a mathematical model to predict urban building energy consumption. One of the advantages of the statistical approach is the ability to predict energy consumption at large scale without the need for significant inputs and assumptions as in engineering-based method. Statistical models, however, are in general, providing less detail and flexibility when it comes to evaluating the impact of a wide range of energy conservation scenarios of new technologies.

The bottom-up engineering method often requires more detailed building characteristics data as model inputs, for instance, occupancy profile, thermostat setting, air infiltration rate, and so on (for the concerned parameters in this study see Tables 3 and 4), and it simulates energy demand based on the science of building physics. The capability to generalize and predict system behaviour given previously unobserved conditions is the most significant advantage of this modelling approach. Due to those characteristics, it is a versatile approach to assist decision-making and scenario analysis. For instance, it has potentials to be used in evaluating (urban) building energy performance, identifying cost-effective building retrofit measures, quantifying the impacts of future climate scenarios on energy consumption and effectiveness of adopting new technologies, assessing energy demand and supply balance, and supporting energy-efficient building design and district planning (Swan and Ugursal 2009; Kavgic et al. 2010; Allegrini et al. 2015).

Although a bottom-up method is a more comprehensive tool to assess dynamic energy consumption, the high computational cost of applying a building energy simulation engine at the urban scale is a significant challenge. Another obvious obstacle is how to properly deal with inherent simulation uncertainties. Particularly, input uncertainties either caused by subjective uncertainty or stochastic uncertainty could be significant factors to cause simulation performance gaps (Keirstead, Jennings, and Sivakumar 2012; Coakley, Raftery, and Keane 2014; Fonseca and Schlueter 2015; Reinhart and Davila 2016; Nouvel et al. 2017). In the first case, a single-valued parameter does exist, but it is unknown to the modeller because of incomplete information. For instance, the thermal transmittance of the construction material, namely, *U*-values of walls, roofs, floors, and so on. In addition, some simulation inputs are inherently uncertain and fluctuating. For instance, it makes little sense to define occupancy schedule, thermostat setting of the building with fixed assumptions, although this is indeed done in practice very often due to lack of available data.

1.1. Model calibration as a means of bridging the gap

To minimize input uncertainties and simulation performance gaps, model calibration and optimization have been extensively studied and applied at building energy simulation (BES). At BES level calibration, the Bayesian approach (Kennedy and O'Hagan 2001) is proposed and attempts to infer the best-fitting parameter values from the posterior distributions of uncertain parameters. The subjective, uncertain parameters could potentially be reduced to a single value if sufficient data is available; while the distribution of the stochastic type uncertain parameter could be effectively refined to describe the underlying random process, thus reducing the discrepancy between the model prediction and observed energy data (Heo 2011; Booth, Choudhary, and Spiegelhalter 2012; Coakley, Raftery, and Keane 2014; Machairas, Tsangrassoulis, and Axarli 2014).

Nonetheless, bottom-up engineering modelling and calibration at the urban scale remains a significant challenge as it is way more expensive to acquire detailed building data at scales or measured energy consumption data required for validation and calibration is usually incomplete or at aggregated levels due to the constrained of the current registration system or privacy concerns. (Keirstead, Jennings, and Sivakumar 2012; Reinhart and Davila 2016).

Archetype modelling, which seeks to reduce the number of buildings via segmenting the building stock into a smaller subset of homogeneous archetypes (energy representative of sample buildings), seems to be a plausible solution. Among 17 works reviewed by Reinhart and Davila (2016), the number of archetypes varies from 5 to 3168. In many studies, building usage type, construction year, geometry, floor area, and etc., are some commonly used classifiers (Guerra Santin, Itard, and Visscher 2009; Howard et al. 2012; Mastrucci et al. 2014; Nouvel et al. 2015; Torabi Moghadam et al. 2018). The TABULA (Loga, Diefenbach, and Stein 2012) and the follow-up EPISCOPE[1] project are notable efforts to develop archetypes among 21 European nations. Monteiro et al. (2018) partition partial residential building stock of Lisbon into 18 archetypes based on construction period, size class (single-family, multi-family), roof type, and neighbouring. Ghiassi and Mahdavi (2017) investigate three multivariate cluster analysis (MCA) methods, which are hierarchical agglomerative clustering, K-means clustering, and model-based clustering, respectively, and each method generates different numbers of clusters under different scenarios.

Although archetype modelling potentially reduces the requirements of individual building data acquisition. The remained challenge is that there is no general archetype definition because fundamental interactions may differ according to local circumstances. Consequently, the process of classifying and characterizing building archetype often relies on empirical assumptions, expert judgments, and examples from the literature, which can involve many uncertainties and lead to erroneous results in some occasions. Besides, classifying the building stock based on a few parameters potentially leads to a loss in the natural variability when information that would allow further differentiation is limited or unknown.

Using probabilistic modelling and calibration instead may be a reasonable approach to incorporate archetype heterogeneity and parameter uncertainties more appropriately, however, there are only a few attempts to do so. Booth, Choudhary, and Spiegelhalter (2012) integrate a probabilistic sensitivity analysis with a Bayesian calibration framework (Kennedy and O'Hagan 2001) based on a monthly average quasi-steady-state model and calibrate uncertain parameters of a group of 35 physically similar flats using metered electricity data over 61 consecutive winter days. Cerezo et al. (2015), Cerezo et al. (2017) and Sokol, Cerezo Davila, and Reinhart (2017) iteratively segment the residential building stock of Kuwait City and Cambridge, Massachusetts respectively into different numbers of archetypes with increasing levels of details. Among the most detailed archetype definition, the selected occupant-related uncertain parameters are further modelled in a probabilistic

way. The model is based on dynamic simulation engine EnergyPlus and calibrated with monthly or annual metered energy data of the same year. A threshold is defined for the simulated errors as a binary likelihood function to filter building-specific calibration parameters for individual building independently. The inferred occupant-related uncertain parameters are subsequently merged into archetype specific posterior joint distributions for prediction.

Model calibration is an over-specified and under-determined problem. In most cases, there are many model inputs but comparatively few energy measurements available. This could lead to over-fitting issues (Kennedy and O'Hagan 2001; Coakley, Raftery, and Keane 2014). To mitigate this outcome and have a valid calibration, sensitivity and uncertainty analysis (Saltelli et al. 2008; Wei 2013) can assist in identifying the most influential variables affecting the simulation results. The modeller can thus prioritize data collection and calibration targets or make more sensible assumptions for the prior probability distributions of these key uncertain inputs.

1.2. Contributions of the paper

Most existing bottom-up engineering urban energy models have relied on explicit parameter choices for the building stock due to limited data availability, and only a few studies have gone through a calibration process. Simulation performance gaps are expected if the provided inputs cannot accurately depict the corresponding buildings and building blocks.

Based on the archetype probabilistic modelling and calibration framework proposed by Cerezo et al. (2015), Cerezo et al. (2017) and Sokol, Cerezo Davila, and Reinhart (2017), we further incorporate sensitivity analysis to select the influential uncertain variables at the corresponding modelling resolution, which can be either stochastic or subjective uncertainties, instead of exclusively calibrating occupant-related parameters based on the empirical assumption. The heating demand model is based on a dynamic urban energy model: CitySim (Robinson et al. 2009) and models each building as a single thermal zone and aggregate to postcode level (tens of households). Preparation and simulation efforts of using this energy model are discussed in the text. Instead of using metered data of the same year, this work expands the training data with six consecutive years of annual gas consumption data and validates the results with two subsequent years of measurement data to check the validity of the method when new environmental and weather conditions applied. Rather than using defined simulated errors as a binary likelihood function to filter building-specific calibration parameters proposed by Cerezo et al. (2015), Cerezo et al. (2017) and Sokol, Cerezo Davila, and Reinhart (2017), this work demonstrates that significant improvements in model accuracy can be obtained even using simple uncertainty models, i.e. normal distributions with variances estimated from the data.

Data availability is a crucial barrier for such analysis, requiring measurements that span multiple years and cover a large geographic area for which building models are available. The study performed in this paper is limited to postcode-6 resolution (tens of households) and annual gas consumption for this reason, although the methodology is applicable to data sets with higher resolution. Collecting and harmonizing heterogeneous local data to perform urban scale energy modelling is often regarded as a challenging and time-consuming task, this work provides source references and example procedures to manage open-source data into a valid 3D city model for urban building energy modelling which is applicable to most cities in the Netherlands.

2. Energy model and simulation inputs

2.1. Urban energy model: CitySim

Among numerous simulation tools, CitySim is adopted in our research considering the following characteristics. CitySim models the dynamic irradiation on the exterior building surfaces to consider the effects of inter-building obstruction. In addition, a resistor-capacitor (R-C) thermal model is implemented in CitySim to calculate the thermal exchange between the outdoor and indoor environment (Robinson et al. 2009). As a consequence, CitySim can simultaneously consider important geometric features at an urban scale, including the building size, shape, orientation and density in response to local weather data and at the appropriate level of detail. The simplified building thermal model not only reduces computational cost but also eases the burden of managing detailed building level data, which is often the biggest obstacle at an urban scale simulation.

CitySim requires 3D building models and ground surface model to simulate environmental interactions in the built environment. Building geometry and construction details such as facade *U*-values (thermal transmittance coefficient [W/m^2K]) and window to wall ratios (glazing ratio) of each surface, operation details such as the number of occupants and the occupancy profile of the building can be specified in the CitySim XML format or OGC (Open Geospatial Consortium) CityGML format (Gröger et al. 2012).

2.2. Data preparation and uncertainty quantification

While the required inputs for running an energy model are significantly dependent on the adopted simulation engine and modelling purpose, they can, in general, be grouped into the following data categories: weather, geometry, construction, energy system, operation, and energy consumption. These data categories also apply to CitySim inputs. Tables 3 and 4 summarize the investigated and modified CitySim parameters in this case study.

In the final step of data preparation, all heterogeneous datasets are cleaned, harmonized and integrated mostly by GIS operations (e.g. FME Desktop[2]) and Python scripts. The end results are managed in the PostgreSQL database and some as GIS layers in shapefile format (Environmental Systems Research Institute 1998).

2.2.1. Weather data

Historical observation records measured at Schiphol weather station (approximately 10 km from the city centre of Amsterdam) are accessed from the Royal Dutch Meteorological Institute (Koninklijk Nederlands Meteorologisch Instituut, KNMI.[3]) Also, diffuse horizontal irradiance values and solar normal irradiance

values are supplemented from EnergyPlus weather data repository, Amsterdam EPW.[4] Overall, 8 years of meteorological data from the year 2010 to 2017 are collected and translated into CitySim compatible weather files.

2.2.2. Building geometry

In the Netherlands, the Basic Registration Addresses and Buildings data (Basisregistratie Adressen en Gebouwen, BAG) managed by Kadaster[5] contains detailed, up to date, and georeferenced building (BAG.pand) and address (BAG.verblijfsobject) data of the entire country. Attributes like the year of construction, building function and building footprint, and etc., are included in BAG and can be freely accessed via Web Feature Service (WFS) which is maintained by Nationaal Georegister (NGR).[6]

Building geometry is modelled as a level of detail 1 (LOD1) block model, the coarsest volumetric representation defined in the Open Geospatial Consortium (OGC) CityGML standard (Gröger et al. 2012). A LOD1 building model is usually acquired with extrusion from 2D building footprint with building height estimated from point cloud data (Ledoux and Meijers 2011). 2D building footprint in the current implementation is a standard GIS file in shapefile format (Environmental Systems Research Institute 1998). Point cloud data (Actueel Hoogtebestand Nederland, AHN3) provides detailed and precise elevation data collected by airborne laser scanning techniques (or LiDAR: Light Detection And Ranging) and has an average of eight height points per square meter covering the whole Netherlands. These open datasets are accessed from PDOK.[7] As building reference height is estimated from AHN3 point cloud data (median height of the points positioned within the building footprint). We assume the building reference height has uncertainty range of 90% to 110% of the estimated height to account for the building height estimation uncertainty caused by different roof types. At the time of writing, Dukai, Ledoux, and Stoter (2019) further investigate building height uncertainty estimation of the Netherlands and conclude that LOD1 building geometry generated based on this method is suitable for most GIS-related analyses.

The level of detail of the collected data is mostly at building and postcode scale. To reduce simulation complexity, each building in this work is modelled as a single thermal zone. This dramatically simplifies geometric processing complexity, even though defining multiple thermal zones within a building is possible in CitySim. To generate a valid 3D city model for thermal simulation in CitySim, walls between adjacent buildings were removed, to prevent CitySim from overestimating heat losses to the external environment.

2.2.3. Construction data

Construction data refer to the thermal transmittance coefficient (*U*-value) of roofs, walls, floors and windows (denoted as Uroof, Uwall, Ufloor, Uwindow); solar energy transmittance of window glazing (Gwindow); building infiltration rate (Ninf); window to wall ratio, window to roof ratio (WWR, WRR); (ground) surface shortwave reflectance (GSW, SW). While it is impractical to collect specific construction parameters for each building, some construction parameters, especially *U*-values, are related to and can be inferred from the building construction periods. These building year-dependent data can be found from sources such as the European Building Stock Observatory[8], EPISCOPE and TABULA project web portals[9], Ecofys report (Petersdorff et al. 2005), Sociale Huursector Audit en Evaluatie van Resultaten Energiebesparing (SHAERE) database (Filippidou 2018). The uncertainty range of each of the thermal transmittance coefficient (*U*-value) is derived based on the minimum and maximum values found in three references: (Petersdorff et al. 2005; Loga, Diefenbach, and Stein 2012; European Commission 2015).

At the stage of data collection, we experienced that *U*-value references are generally more accessible and comprehensive than the other construction parameters. This could be due to high data collection cost (e.g. building infiltration rate), or the parameter itself has stochastic variability and therefore is not readily characterized by a single value. Considering these data limitations, uncertainty ranges for parameters WRR, Gwindow, SW, GSW are made based on generic assumptions and presented in Table 4.

The parameter Ninf stands for building air volume change rate per hour in normal conditions and takes air infiltration through the envelope, airflow through the doors, windows, and casual or for ventilation purposes into account (Perez 2014). Due to the measurement complexity, this is an influential but one of the least accessible parameters. The uncertainty range is thus mainly based on the discussion via (Perez 2014) and cross-referenced with the other studies (Alfano et al. 2012; Bramiana, Entrop, and Halman 2016).

2.2.4. System data

Although the share of natural gas-powered heating system is decreasing, the majority of heating demand (80%) is fulfilled by the combustion of natural gas through boilers or cogeneration plants in the Netherlands (Energy Research Centre of the Netherlands 2015). Table 1 serves as a reference for the share of household heating system type and efficiency of the non-profit buildings accessed from SHAERE database (Filippidou 2018). According to this summary table, we adopted 0.80 and 0.95 as a lower and upper bound of the uncertainty range of heating system efficiency (average per postcode area). Meanwhile, the table clearly shows that the condensing boiler with an efficiency $\eta \geq 0.95$ is the dominant installation in the existing building stock. As a consequence, 0.95 is adopted as a baseline value. Also, accessing detailed heating system type and system efficiency per building or even postcode level is challenging due to privacy concerns. As a result, in the current implementation, heating system diversity and uncertainty is treated as homogeneously distributed in the building stock when compared with simulation inputs.

2.2.5. Operation data

In the context of this research, operation parameters refer to the number of occupants per building, occupancy schedule (profile), minimum thermostat setting (Tmin), and window openable ratio (WOR). Occupant per residential building is derived from postcode statistics published by the Central Bureau of Statistics (Centraal Bureau voor de Statistiek, CBS[10]) of years 2008 to 2010.

The minimum thermostat setting of the heating system is another influential yet uncertain input on heating demand calculation. In the studies (Leidelmeijer and van Grieken 2005; Guerra Santin, Itard, and Visscher 2009), which are partially based

Table 1. Distribution of heating system types in the Netherlands. Retrieved from Filippidou (2018).

Type of heating system	Frequency	Percentage
Condensing boiler ($\eta \geq 0.95$)	930,127	73.7
Improved non-condensing boiler ($\eta = 0.80 - 0.90$)	178,557	14.2
Condensing boiler ($\eta = 0.90 - 0.925$)	42,026	3.3
Gas/oil stove	40548	3.2
Conventional boiler ($\eta < 0.80$)	29,973	2.4
Condensing boiler ($\eta = 0.925 - 0.95$)	19,595	1.6
Heat pump	16,722	1.3
μCHP	2751	0.2
Electric stove	484	≈ 0.0
Total	1,260,783	100.0

Table 2. Thermostat setting (°C) profiles in Dutch households. Adapted from Leidelmeijer and van Grieken (2005).

	Morning	Afternoon	Evening	Weekend	Share
Profile 1	15.2	17.4	14.1	16.2	4%
Profile 2	18.4	18.8	15.6	18.5	16%
Profile 3	19.7	20.2	15.2	20.0	35%
Profile 4	19.6	20.0	11.6	19.8	8%
Profile 5	14.9	20.2	14.7	20.1	11%
Profile 6	20.9	21.2	20.4	21.1	5%
Profile 7	21.6	22.0	15.5	21.7	20%

on the WoON survey[11] data in the Netherlands, several different temperature setting profiles throughout the different time of a day and weekend are observed. These are summarized in Table 2. The weighted average value 18.38°C is adopted as a baseline input. This profile also gives an insight and helps the modeller to quantify uncertainty range of the minimum thermostat setting in average Dutch households by taking the minimum and maximum temperature as lower and upper bounds from the weighted average (approximately between 15 °C and 20 °C).

Window openable ratio (WOR) is rarely mentioned in technical report and literature reviewed by the authors so that the baseline value and uncertainty range of this parameter are based on generic assumptions as presented in Table 4.

2.2.6. Energy consumption data

Metered gas consumption data is not only used to validate simulation results but also applied to model calibration. Ten years of annual energy consumption records from 2008 to 2017 at postcode level are made available via the Liander (distribution system operator) open data portal.[12] To our knowledge, this is the smallest spatial-temporal resolution energy data of the study area made publicly available by the time of conducting this research. The energy consumption records also come with detailed metadata, such as delivery percentage (network supply minus customer self-generation), network status, and so on, which gives the user better control over data quality.

2.3. Input summary

Based on non-exhaustive literature and technical reports, baseline values and the corresponding uncertainty ranges of the simulation parameters are summarized in Tables 3 and 4 as listed below.

2.4. Heterogeneous datasets integration and semantic 3D city model enrichment

The entire data harmonization and 3D city model enrichment process involves several steps. It is accomplished by combining multiple tools (environments) in use, such as GIS processing in FME software, geometry processing in Rhinoceros 3D software together with Grasshopper plugin, PostgreSQL database and Python script.

After the data filtering and cleaning process, heterogeneous spatial datasets are harmonized into a GIS layer: integrated baselayer (see Figure 1 for the data model), while non-spatial datasets are managed in a database. As the CitySim 3D city model is managed in XML format, it can be easily parsed and overwritten by the developed Python script. The main task done by the Python script is to retrieve building information from the integrated baselayer and database and to overwrite the CitySim default values with the baseline values or probabilistically sampled for each building.

3. Methodology

3.1. Implementation

In this case study, 2178 buildings are included in the area of interest (Figure 2). Considering the resolution of the best available data, all collected and cleaned open-source datasets mentioned in the previous section are aggregated at the postcode level rather than an individual building. At least 84 residential postcodes fulfil simulation data requirements (see Figure 1) and will thus be used for the model calibration process.

Because of incomplete floor area information occasionally found in the GIS layer, building (postcode) volume can be more accurately estimated than floor area per building (postcode) when a 3D city model is available. As a consequence, contrary to the prevailing convention, the Energy Use Intensity (EUI, kWh/m^3) unit is normalized over cubic meter instead of squared meter.

3.2. Sensitivity analysis

Depending on the modelling purposes, energy simulation can sometimes be very sophisticated and requires tens to hundreds of inputs to run (Swan and Ugursal 2009; Coakley, Raftery, and Keane 2014; Hsu 2015). Identifying the key parameters influencing heating demand calculation can be a critical step to have an effective calibration result, especially when available datasets and computational resources are constrained.

The Morris method is adopted due to its capability to give parameter importance ranking with fewer computational resources (Saltelli et al. 2008). The Morris method is a global sensitivity method to evaluate the influence of uncertain parameters over the whole parameter range. The sensitivity analysis is carried out by the Python script and external library, SALib[13] (Herman and Usher 2017), which includes commonly used sensitivity analysis methods.

A stand-alone cubic building with 13.5 m in all dimensions positioned in the centre of the ground surface is set up for the

Table 3. Data collections applied to the Amsterdam UBEM development.

Data category	Dataset	Data period	Remark	Source
Weather	Annual hourly observation	2010–2017	[a]	KNMI
	Irradiance	1982–1999	[b]	Amsterdam.epw
Geometry	Building footprint	up to date	–	BAG (WFS)
	Building height	2014–2019	[c]	AHN3
System	Heating system statistics	2010–2014	[d]	Filippidou (2018)
Operation	PC6 population	2008–2010	[e]	CBS
	Occupancy schedule	–	[f]	ASHRAE
	Temperature set-point	– 2005	–	Leidelmeijer and van Grieken (2005)
	Window openable ratio	–	[g]	–
Construction	See Table 4	–	–	–
Energy	PC6 annual gas consumption	2010–2017	[h]	Liander

[a]The measurement is made at Schiphol meteorological station, which is approximately 10 km away from Amsterdam centre.
[b]This data is based on IWEC data in Amsterdam and managed in EPW (EnergyPlus Weather) format. IWEC data is derived from long-term observation sometimes up to 18 years (1982–1999 for most stations). Details refer to https://energyplus.net/weather/sources#IWEC.
[c]AHN3 point cloud data collection period is scheduled to 2019 Details refer to http://ahn.maps.arcgis.com/apps/Cascade/index.html?appid = 75245be5e0384d47856d2b912fc1b7ed.
[d]Statistical distribution data of the heating system type and efficiency collected from the non-profit building stock database (SHAERE) in the Netherlands.
[e]Only 2008–2010 is made freely accessible and 2012–2014 data is made available at cost.
[f]Standardized residential occupancy profile.
[g]Barely found reliable data source, rational assumption is made for this parameter.
[h]Liander energy data is better than the CBS data quantitatively and qualitatively for this purpose as it contains several years of consumption data and detailed metadata.

Table 4. Defined baseline values and uncertainty ranges of the simulation inputs.

Parameters	Symbol	Unit	Baseline	Uncertainty	Source
Building construction parameters					
Window to wall ratio	WWR	–	0.21	U(0.15–0.45)	Petersdorff et al. (2005)
Window to roof ratio	WRR	–	0.00	U(0.00–0.15)	–
Thermal transmittance coefficient of roof	Uroof	W/m^2K	Table 5	U(0.16–2.60)	Petersdorff et al. (2005), Loga, Diefenbach, and Stein (2012) and European Commission (2015)
Thermal transmittance coefficient of wall	Uwall	W/m^2K	Table 5	U(0.21–2.55)	Petersdorff et al. (2005), Loga, Diefenbach, and Stein (2012) and European Commission (2015)
Thermal transmittance coefficient of floor	Ufloor	W/m^2K	Table 5	U(0.27–2.09)	Petersdorff et al. (2005), Loga, Diefenbach, and Stein (2012) and European Commission (2015)
Thermal transmittance coefficient of window	Uwindow	W/m^2K	Table 5	U(1.68–3.80)	Petersdorff et al. (2005), Loga, Diefenbach, and Stein (2012) and European Commission (2015)
Solar energy transmittance of window glazing	Gwindow	–	0.76	U(0.30–0.85)	–
Surface shortwave reflectance	SW	–	0.20	U(0.20–0.50)	–
Ground surface shortwave reflectance	GSW	–	0.20	U(0.20–0.50)	–
Infiltration rate (air change rate)	Ninf	Volume/h	0.60	U(0.19–0.81)	Alfano et al. (2012), Perez (2014) and Bramiana, Entrop, and Halman (2016).
Operation parameters					
Minimum thermostat setting	Tmin	°C	18.38	U(15.0–20.0)	Leidelmeijer and van Grieken (2005).
Window openable ratio	WOR	–	0.25	U(0.00–0.35)	–
System parameter					
Heating system efficiency	Eta	–	0.95	U(0.80–0.95)	Filippidou (2018).
Geometry parameter					
Building height uncertainty	B_h	–	–	U(0.90–1.10)	Biljecki, Ledoux, and Stoter (2014)

sensitivity analysis. Input uncertainty ranges used in the sensitivity analysis are summarized in Table 4. All 14 inputs (k) with the corresponding uncertainty ranges are normalized to scale [0, 1] and divided into 10 levels (p) respectively, which leads to 10^{14} input combinations (Ω). The grid jump size (Δ) is set to be 2. The Morris method sequentially steps along each of the k parameters, thus generating trajectories of $k+1$ points in the input space (Ω). A Monte Carlo approach is used to generate t independent trajectories, for a total number of evaluations $N = t \times (k+1)$. An increase in t results in more stable parameter ranking. Different trajectory numbers ($t = 10, 30, 50, 100, 150$) are tested to ensure the stability of the results. The experiment reveals that when $t \geqslant 50$, namely $N \geqslant 750$, is sufficient to give a stable parameter ranking for the further analysis in this case. The procedure for conducting the Morris method is presented in Figure 3.

Table 5. Baseline U-values in different construction periods (Loga, Diefenbach, and Stein 2012).

Parameters	Pre 1965	1965–1974	1975–1991	1992–2005	2006–2014	Post-2014
Uroof	1.68	0.89	0.64	0.36	0.23	0.16
Uwall	1.76	1.45	0.64	0.36	0.27	0.21
Ufloor	1.75	2.09	0.94	0.35	0.27	0.27
Uwindow	2.90	2.90	2.90	1.80	1.80	1.80

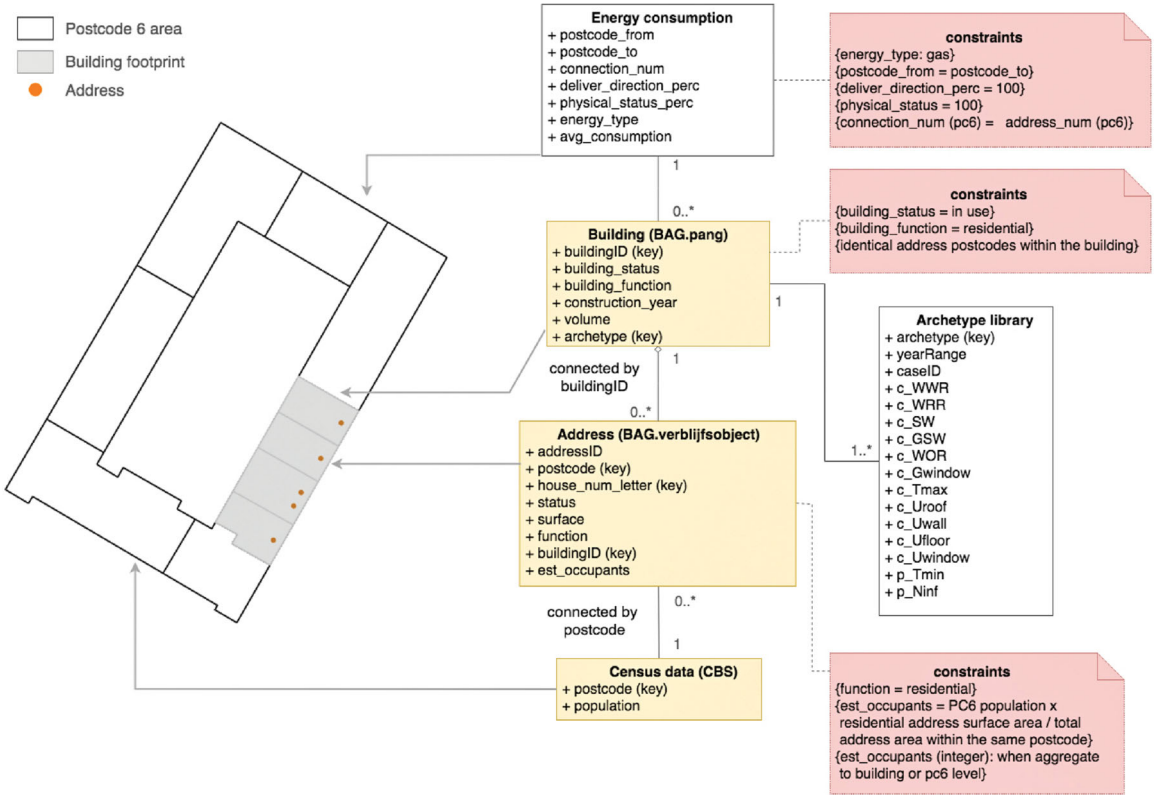

Figure 1. The UML diagram shows the data model and data filtering criteria in this study. The spatial resolution of each data is illustrated on the left. The integrated baselayer is harmonized from different source layers (Building, Address, and Census data), and the data filtering rules are indicated with the constraints headers on the right.

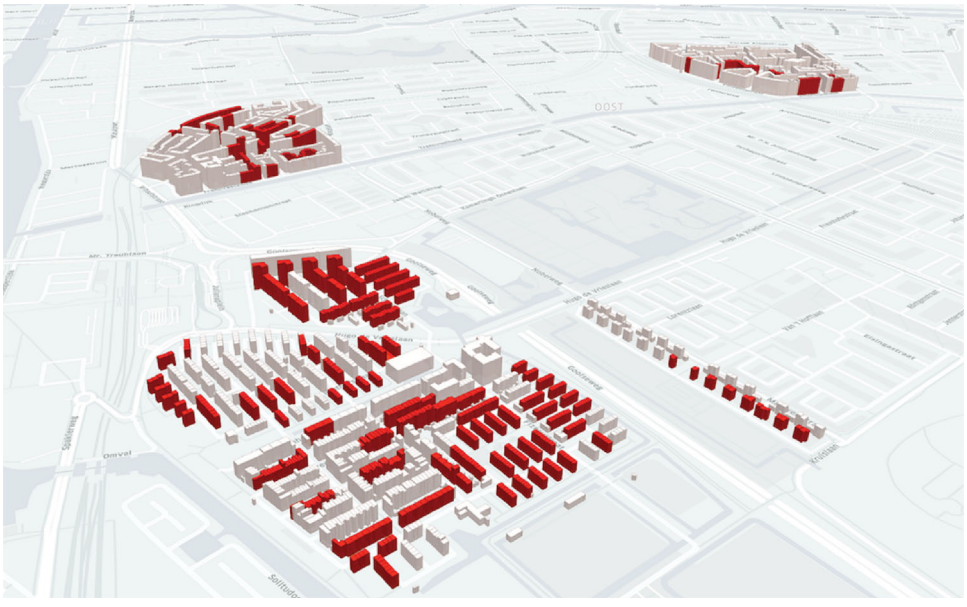

Figure 2. 3D city model of the partial districts in Amsterdam-Oost. Buildings (postcodes) coloured in dark have sufficient data to perform heating demand simulation and calibration.

With the setting mentioned above, the operation parameter *Tmin* (minimum thermostat setting) is found to have the most significant influence on annual heating demand calculation, followed by construction parameters *Uwall* and *Ninf* . Furthermore, the result presented in Figure 4 clearly shows that construction parameters, especially *U*-values of the wall, floor, window, and roof have significant to moderate influence on annual heating demand calculation. Meanwhile, building height (B_h), building surface shortwave reflectance (SW), window to roof ratio (WRR), ground surface shortwave reflectance (GSW), and window openable ratio (WOR) are insignificant. The ignorable result of SW and GSW might be caused by the experimental setting, where no surrounding buildings are presented.

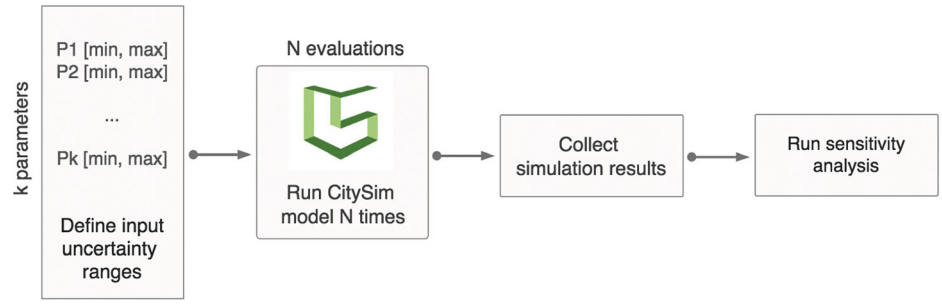

Figure 3. The procedure to perform Morris sensitivity analysis in this study.

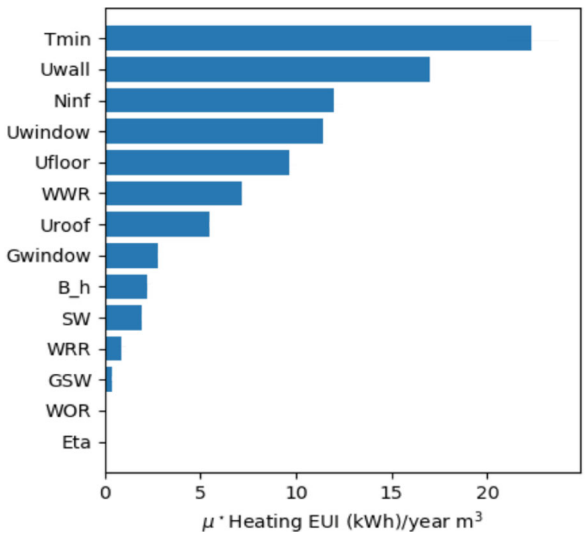

Figure 4. Sensitivity analysis result based on a simple cubic building with $k = 14$, $p = 10$, $\Delta = 2$, and $t = 50$. Diagram shows parameter importance ranking based the average elementary effect μ^*.

Further investigations revealed that in this sensitivity analysis based on CitySim environment, simply providing system efficiency (Eta) alone cannot effectively calculate the associated influence on heating EUI when another CitySim module input, the maximum heating thermal power, is not adjusted accordingly. Nevertheless, it is challenging to collect detailed information of maximum thermal power of the heating system at this spatial level, and thus prior knowledge of this parameter remains ambiguous. As a consequence, this sensitivity analysis result is interpreted based on 13 useful parameters in the end.

3.3. Postcode probabilistic modelling

The building stock of the studied area is classified into 18 archetypes based on Dutch national reference home standard (Agentschap NL – Ministerie van Binnenlandse Zaken en Koninkrijksrelaties 2011, 2013) and EPISCOPE and TABULA project (Loga, Diefenbach, and Stein 2012) as we can associate explicit energy-relevant building parameters, e.g. associating building construction *U*-values based on the building construction period, from the corresponding database. Based on this classification scheme, each residential postcode in this case study is assigned with one archetype definition based on its averaged construction period and dwelling type, see Figure 5. While the Dutch standard and TABULA have six and four dwelling types respectively, they are aggregated into three main types in this research: single-family house (SFH), terrace house (TH) and multi-family house (MFH).

According to the sensitivity analysis result, calibrating the key uncertain parameters is most likely to minimize heating demand simulation performance gaps. As a consequence, two key uncertain parameters (*Tmin* and *Ninf*) are selected, and the corresponding uncertainty ranges are divided into 5 sections, respectively, with an uninformative prior distribution assumption (uniform probabilities). These prior distributions are assigned to individual postcode and will be calibrated independently in the next step. We note that this implicitly assumes a single building model (and unique values of Tmin and Ninf) per postcode area. When more data is available for calibration, a more fine-grained approach could be taken.

We choose the first (*Tmin*) and the third parameters (*Ninf*) for calibration instead of the second one (*Uwall*) because Uwall, Uroof, Ufloor are often bound together and correlated to the construction year; or alternatively dependent on the effective year built, if the building has undergone the major building retrofit. This indicates calibrating Uwall alone is likely to lead to an over-fitting result if not taking Uroof, Ufloor into account, while calibrating additional Uroof, Ufloor will lead to a substantial increase in complexity. This is under consideration for the future works.

3.4. Bayesian calibration framework

All of the previous steps lead to two important outputs, explicit characterization of the building stock according to the assigned archetype, which is grouped into a vector $\boldsymbol{x}_p$, and two key uncertain parameters, *Tmin* and *Ninf*, denoted as a random vector $\boldsymbol{\Theta}_p$, where the index p refers to the postcode. *Tmin* and *Ninf* are each divided into 5 levels with a uniform prior probability distribution respectively, and this leads to totally 25 input combinations. Together, $(\boldsymbol{x}_p, \boldsymbol{\Theta}_p)$ thus can be interpreted to generate 25 models for each valid postcode. We are particularly interested in which input combination, $\boldsymbol{\theta}_p$, is most likely to be correct, given the simulation model and the metered data E_{py}, where y is the measurement year in the training set. This is determined using the posterior probability $P(\boldsymbol{\theta}_p|E_{py})$, calculated according to Bayes' theorem (1), where $P(E_{py}|\boldsymbol{\theta}_p)$ is the likelihood function and $P(\boldsymbol{\theta}_p)$ is prior probability.

$$P(\boldsymbol{\theta}_p|E_{py}) = \frac{P(E_{py}|\boldsymbol{\theta}_p)P(\boldsymbol{\theta}_p)}{P(E_{py})} \tag{1}$$

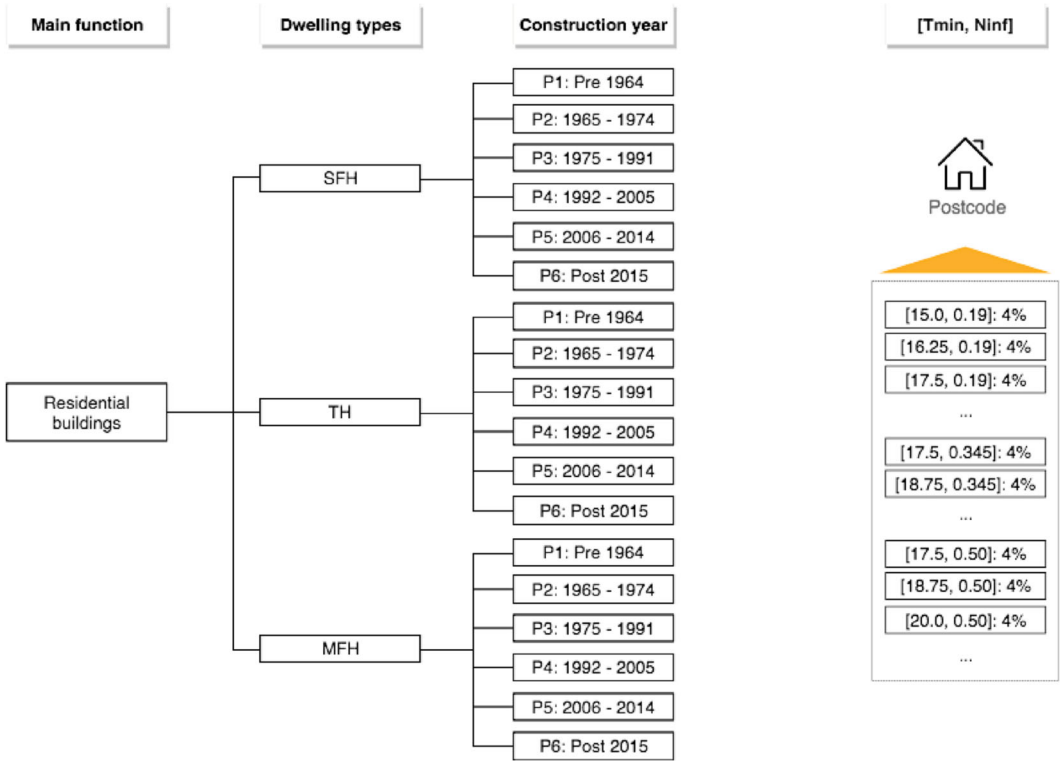

Figure 5. The chart on the left shows the adopted archetype classification rules. The chart on the right shows each postcode is independently assigned with uninformative prior assumptions of *Tmin* and *Ninf*.

and

$$P(E_{py}) = \sum_{\theta_{Ninf}} \sum_{\theta_{Tmin}} P(E_{py}|\boldsymbol{\theta}_p) \times P(\boldsymbol{\theta}_p) \tag{2}$$

In reality, there are many (independent or correlated) factors that can affect likelihood function, and the explicit form does not exist. As a consequence, we assume the likelihood function $P(E_{py}|\boldsymbol{\theta}_p)$ can be described by a Gaussian normal distribution function as shown below.

$$P(E_{py}|\boldsymbol{\theta}_p) \approx P(E_{py}; \mu_{py}, \sigma_p) = \frac{1}{\sigma_p\sqrt{2\pi}} \exp\left(-\frac{(E_{py} - \mu_{py})^2}{2\sigma_p^2}\right) \tag{3}$$

where E_{py} is the measured EUI (kWh/m^3) of the individual postcode p of the training year y; μ_{py} is the simulated EUI of the corresponding postcode of the same year given the specific input combination $\boldsymbol{\theta}_p$; the standard deviation σ_p accounts for the inherent variability of energy consumption in the postcode. We estimate the value of σ_p as 6.8% of $\overline{E}_p$, where $\overline{E}_p$ is the average EUI of the individual postcode over the six training years. This estimate follows from a least squares fit of the linear model (component-wise)

$$e_{py} = \sum_{q \in \text{postcodes}} a_q \delta_{qp} + \sum_{z \in \text{years}} b_z \delta_{zy} + n_{py} \tag{4}$$

where the dependent variable, $e_{py} = E_{py}/\overline{E}_p$, is the normalized postcode EUI and a_q and b_z are fitting coefficients per postcode q and year z. The postcode and year are used as features with 'one-hot' encoding via the Kronecker deltas ($\delta_{ij} = 1$ if $i = j$, and 0 otherwise). The resulting distribution of the residuals n_{py} is shown in Figure 6. Besides a few high-end outliers, the normalized residual errors collected from the least square fitting approximates to a normal distribution and with a standard deviation ≈ 0.068, see Figure 6. This result gives us a confidence to assume that 6.8% of $\overline{E}_p$ can be used as σ_p in Equation (3) to describe typical EUI variations.

4. Result and discussion

In this case study, we developed an urban building heating demand model with CitySim for the partial districts of Amsterdam based on open-source data collections and the model is calibrated with six years of measured consumption data.

In the training phase, annual metered data from the year 2010 to 2015 are used to train the model (Figure 7). The prior probabilities of $P(\boldsymbol{\Theta}_p)$ are initialized as a uniform distribution, so that each value of $P(\boldsymbol{\theta}_p)$ is equiprobable. An iterative calibration process uses the posterior probability of the N year as a new prior of the $N+1$ year. When the training phase is complete, the input combination $P(\boldsymbol{\theta}_p)$ with the highest posterior probability is selected as a calibrated input for each valid postcode to rerun heating demand simulation with 2016 and 2017 weather data. Validation is performed by comparing the baseline simulation result as well as the calibrated simulation result to the measurement data, using the absolute percentage error defined in Equation (5). Coefficient of variation of the root mean square error (CVRMSE) defined in Equation (6) is also applied to measure how well the model fits the measured values at validation period 2016 and 2017, respectively.

$$PE = \left|\frac{EUI_{metered} - EUI_{sim}}{EUI_{metered}}\right| \times 100\% \tag{5}$$

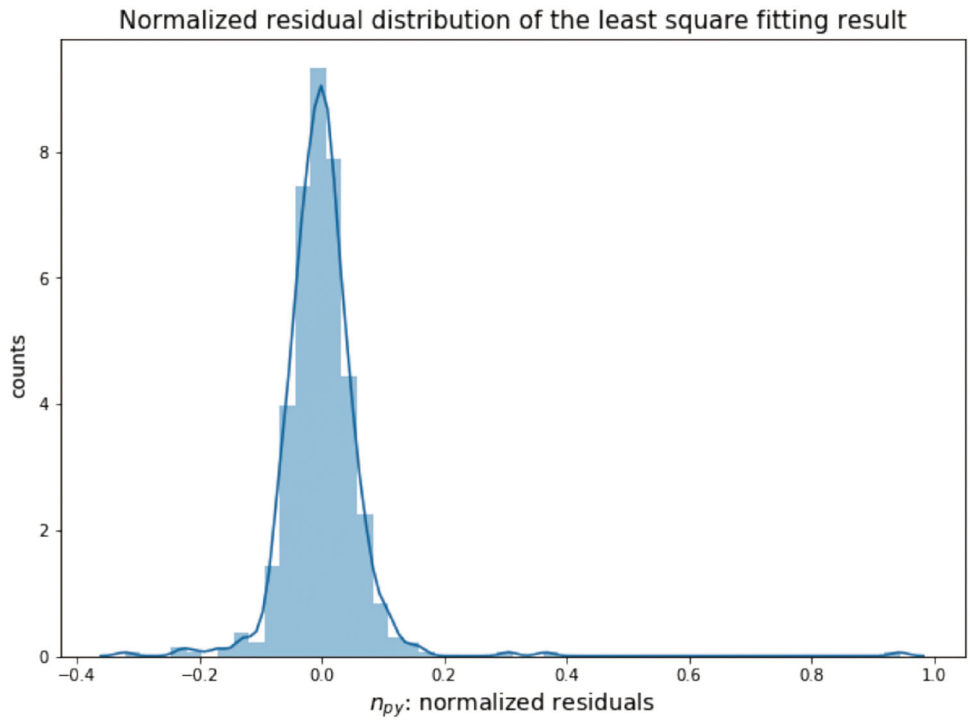

Figure 6. Residual distribution of the normalized EUI data.

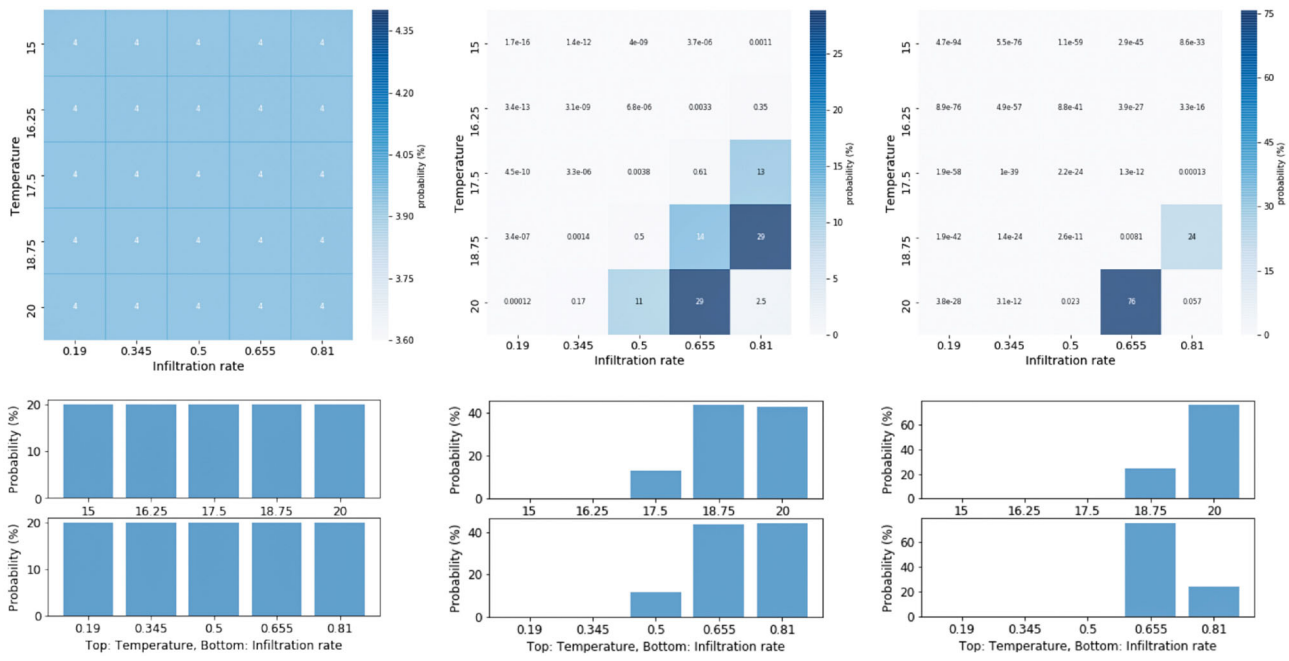

Figure 7. Visualizing the Bayesian inference process of the example postcode 1094 SH throughout the training phase. The joint probability distribution is visualized in a 2D grid in the top, while marginal probability distributions of *Tmin* and *Ninf* are presented in the bottom. From left to right: the prior probability, the posterior after use of the 2010 data, and the posterior after use of data up to and including 2015.

$$CVRMSE = \frac{100}{\bar{y}} \times \sqrt{\frac{\sum_{i=1}^{N_p}(y_i - \hat{y}_i)^2}{N_p}} \tag{6}$$

where N_p is the number of postcode EUI measurements, y_i is the metered EUI for the ith postcode, while $\hat{y}_i$ stands for the simulated EUI, and $\bar{y}$ is the mean of the N_p metered EUI values.

Comparing the baseline and the calibrated simulation results (Figure 8 and Figure 9), the averaged absolute percentage error at postcode level of the validation years, 2016 and 2017, has improved from 25.0% to 8.3% and 19.9% to 7.7% respectively. Meanwhile, postcodes with absolute percentage error less than 10.0% has increased from 23.8% to 75.0% in 2016 (number of calibrated postcodes = 84) and from 32.9% to 78.8% in 2017 (number of calibrated postcodes = 85). Besides, the results show $CVRMSE_{2016} = 11.5\%$ and $CVRMSE_{2017} = 13.2\%$. This indicates that a representative set of parameters was estimated for the calibrated postcodes.

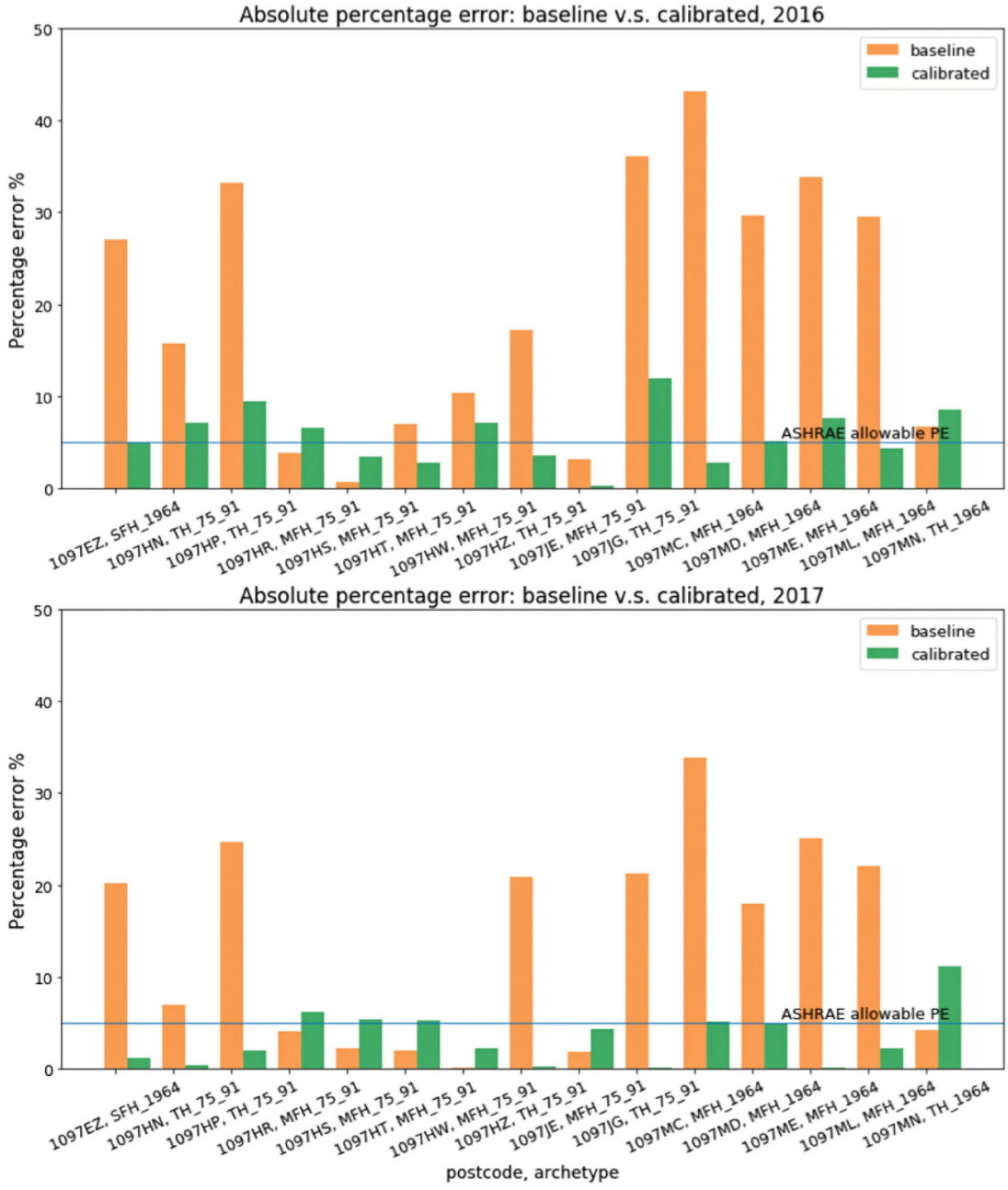

Figure 8. Baseline and calibrated simulation results of the partial postcodes presented in absolute percentage error. The horizontal line indicates 5% of allowable percentage error suggested by the ASHRAE standard.

The proposed Bayesian calibration framework suggests that in the absence of specific information about probability distributions, significant improvements in model calibration can be obtained by using uniform prior distributions and assuming the likelihood functions for energy intensity to be Gaussian. The standard deviation σ_p of the Gaussian distribution function (Equation (3)), which accounts for the inherent variability of the postcode EUI, can be derived from 6.8% of $\bar{E}_p$ (average EUI of the individual postcode over the training years). This is based on a distribution analysis of normalized residual errors resulting from the simple least square fitting (Equation (4)), where the dependent variable is the normalized postcode EUIs of the training years and two variables, *postcode* and *year* are adopted as classifiers.

Following the proposed urban scale calibration methodology, the calibrated UBEM shows acceptable error ranges when it is served as guidance to assist urban planning, retrofit measures assessment, or to provide data-driven decision support (Reinhart and Davila 2016). The methodology and probabilistic modelling also help the modeller to depict the diversity of building stock more accurately.

We note it is beneficial to test if the calibrated parameters can also lead to a better simulation performance at the building level or higher temporal granularity (e.g. monthly or weekly). Validating our results at a finer spatial-temporal resolution is however not feasible, because the best available open-source measurement data is at the annual and postcode level. It is also interesting to investigate the calibration performance when adopting

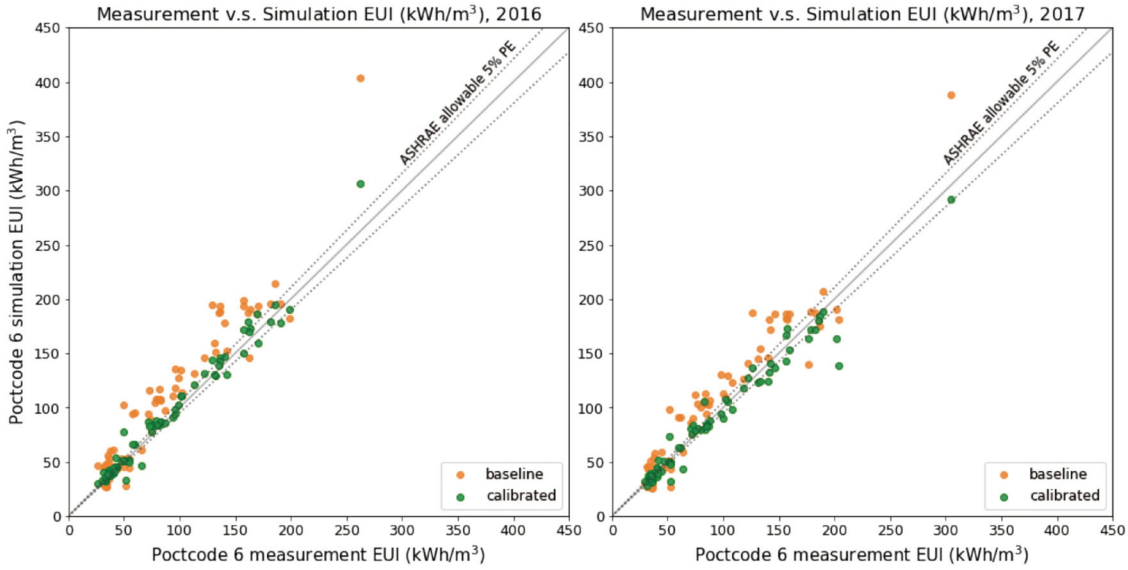

Figure 9. Comparison between the postcode metered gas consumption EUIs and simulation EUIs. The dash lines indicate 5% of allowable percentage error suggested by the ASHRAE standard.

other uncertain parameters (rather than *Tmin* and *Ninf*) from the sensitivity analysis result to perform probabilistic modelling and calibration. Those points mentioned above are under consideration for future work. In the remainder of this section, we discuss distinct possibilities to improve our research.

4.1. Generalizability of the methodology

Constructing and calibrating an urban building energy model is often recognized as a data-hungry process. This study presents an approximate data scope required by this methodology and the efficacy of the calibration results with inputs entirely based on open-source data. In practical situations, data availability and granularity and the critical energy influential parameters vary in different urban areas. Nevertheless, thanks to the crowdsourcing Open Street Map[14], most cities around the world can already be characterized by the primary data (e.g. building footprint, function, or year built). Building height can be estimated if the Digital Surface Model (DSM) and Digital Terrain Model (DTM) are available. Base on these fundamental data sources, further information can be stepwise enriched via GIS processing (e.g. inferring dwelling geometry types) or connecting to the open building (energy) information database, technical reports, and literature to assume the explicit values and uncertainty ranges of the local construction parameters. At a minimum, historical energy consumption data as a training data is required. Through sensitivity analysis, the modeller can determine the most appropriate way to perform location-specific archetype classification and choose the vital uncertain parameters for probabilistic modelling and Bayesian calibration. While there is not yet a standardized approach and it often relies on how the modeller interprets and deals with the variations that could happen in those intermediate steps according to different local circumstances, we expect the methodology to apply to other urban areas as long as the aforementioned fundamental data is available and can be extended to higher data resolutions. The critical energy influential parameters can be identified and collected.

4.2. Data availability and levels of detail

The impact of data availability and granularity on simulation results is evident. As for what kind of spatial and temporal data are suitable for the specific modelling scale and task, sensitivity analysis such as the Morris method provides a very efficient and interpretable way for prioritizing calibration targets. Nevertheless, narrowing down the data collection scope can sometimes be a challenging task already, as the influence of the specific input could be unknown in the early development phase. As a consequence, it is worth to discuss and apply the levels of detail (LODs) concept to energy simulation parameters, according to the modelling domain, spatial and temporal scale of modelling.

This could possibly start with reviewing and summarizing the current works, as many studies have tried to identify the critical variables on different modelling scales and purposes, but no comprehensive review addressing this perspective exists yet. Secondly, sensitivity analysis can be a powerful tool in assisting such a task when high quality and high granularity data is not directly available. When interpreting sensitivity analysis results, one should be cautious that the result is model specific, and the defined uncertainty range has a strong influence on the results. By performing the aforementioned analysis, this could result in a hierarchical LOD framework for energy modelling, which can serve as a guideline for future UBEM development.

4.3. Heterogeneous dataset integration

The open-source data collections used for this case study may not be the best datasets available. For instance, open-source CBS postcode 6 population data is comparatively old, and non-profit building stock databases could be a valuable source if accessibility is authorized. Nevertheless, it is believed that the sources listed in Table 3 provide a good basis for further UBEM development in the Netherlands.

Harmonizing and integrating multi-datasets with different spatial-temporal resolutions is often considered a time-consuming and complex task. Although the data integration

workflow and the specific data model tailored for the project requirement have been made from scratch, to increase data interoperability and to facilitate data exchange for multi-scale and multi-domain simulations, it is worth to consider alternative ways to maintain 3D city models.

An international standardized format such as CityGML, which is based on the Open Geospatial Consortium (OGC) standard (Gröger et al. 2012), and possibly with the support of the Energy ADE (Agugiaro et al. 2018), can be an alternative option. CityGML is based on the Geography Markup Language (GML) to represent and exchange virtual 3D city models. It can define 3D geometry, semantics, ontologies and appearance of most relevant topographic objects of different spatial scales on varying levels of detail (LODs). Besides the existing CityGML thematic modules (bridge, building, city furniture, and so on), it is possible to extend the new classes and attributes by the Application Domain Extension (ADEs) such as Energy ADE, Utility Network ADE (Kutzner and Kolbe 2016), etc., where the Energy ADE is highly relevant to the urban building energy modelling purpose. More future works are required to understand and test how the Energy ADE can support multi-scale and multi-domain simulation and how the standardized data model can fit diverse energy simulation engines and simulation applications.

4.4. Bayesian inference and calibration

It should be reminded that in the proposed Bayesian inference and calibration framework, applying a Gaussian normal distribution to describe the likelihood function, is an assumption as no explicit function exists. The current implementation might fall short to accurately describe the real likelihood distribution for the respective input combinations; for instance, the simulation variability caused by other input uncertainties are not incorporated in this formulation. This is an issue to be addressed in future works.

In a rigorous sense, the parameter posterior distributions should be interpreted on a postcode basis in this study. However, if the parameter posteriors of the specific building archetype (not necessarily confined to the archetype definition of this study) show a statistically significant pattern, the parameter posteriors might be able to apply to the untrained postcodes or buildings to fill the spatial and energy data gap often seen at urban scale modelling.

Computation time is another aspect to be addressed. According to the current implementation, calibrating two parameters requires 25 simulation runs in total, and this process is iteratively conducted 6 times given 6 years of annual gas consumption data. This is 150 simulations per city model in total. Depending on the partitioned city model scale of the test site and geometry complexity (Figure 2), this can require a significant amount of time. A small scale simulation with 226 buildings took approximately 27 h; medium scale with 589 buildings took about 3.75 days; the most extensive scale simulation with 1363 buildings required almost 10 days to complete the expensive training phase on a personal computer with a 4 core 3.60 GHz processor and 16 GB RAM. Although CitySim simulation model scales well in general, such time constraint could still become an obstacle to developing UBEM into an interactive platform for decision support applications.

Table 6. Ten building characteristics are used to calculate definite energy label in the Netherlands (Rijksdienst voor Ondernemend Nederland 2017).

Construction year	Dwelling type
Type of glass	Facade insulation
Room insulation	Floor insulation
Heating system type	Hot water supply type
Ventilation system	Solar panels and solar water heater

4.5. Accelerating building stock retrofit through a calibrated UBEM

One practical use case of the calibrated urban energy model could be to perform extensive scale building performance mapping and labelling. In the Netherlands, registration of a certified energy label when selling, releasing or delivering a house is enforced by law since January 2015. Based on the certified energy label, the house owner can take suggested building retrofit measures if necessary. A provisional energy label, calculated based on publicly registered data such as construction year, etc., is no longer sufficient after January 2015[15] (Rijksdienst voor Ondernemend Nederland 2017). However, registering a definite energy label, which is calculated based on ten building characteristics as listed in Table 6, requires an authorized expert to evaluate an individual house, which is a labour-intensive and time-demanding task. This might explain why more than 50% of buildings did not yet have a definite energy label by the end of 2018[16], thus slowing down the building renovation process. Following the building renovation rates achieved over the years 2010–2014, Filippidou (2018) points out that attaining the short-term goals of upgrading to an average energy label B in the non-profit Dutch housing stock by the end of 2020 is not probable. Based on this fact and the urgent need, calibrated UBEM is a powerful and versatile alternative to perform large-scale building performance mapping and labelling and comes with the capability to carry out retrofit measures assessment and scenario analysis. Also, when the calibrated UBEM is developed into a decision support environment, visualizing energy consumption patterns and retrofit saving potentials could potentially increase citizen engagement, which is one of the critical factors to ensure a successful energy transition.

5. Conclusion

In this paper, we have investigated the urban scale residential heating demand simulation performance gaps caused by input uncertainties and examined the effectiveness of applying the Bayesian calibration approach to resolve this common challenge. The methodology applied in this project has successfully carried out an urban scale heating demand modelling based on a LOD1 3D city model of the mixed-use districts in Amsterdam, constructed entirely from open-source data, and calibrated 84 residential postcodes based on the Bayesian approach, provided with six consecutive years of gas consumption data. The effectiveness of the Bayesian calibration framework is validated when comparing the baseline and the calibrated heating demand simulation results with two additional years of measurement data. Although the model calibration and validation are performed at an aggregated postcode level due to data restriction, we expect the overall methodology is applicable to higher data resolutions.

The code used for the calibration process can be found on the GitHub repository[17].

To ensure an effective model calibration, performing sensitivity analysis is well-advised. The result derived from the efficient and effective Morris method indicates that thermostat setting has the most significant impact on annual heating demand simulation in terms of Amsterdam building stock, followed by building construction parameters such as *U*-values and infiltration rate. Besides, a LOD1 city model should be sufficient to produce acceptable annual heating results based on CitySim.

Modelling the critical uncertain parameters in a probabilistic way can appropriately reflect the imperfect state of knowledge about the urban environment. With the help of Bayesian inference and adequate observation data, parameter uncertainties can be further reduced, and consequently establishing a more reliable UBEM and potentially applying the inferred parameters in other engineering-based models for prediction. Following this framework and adjusting according to the local context, calibrated bottom-up heating demand energy models can be developed in most cities in the world as long as sufficient built environment data is provided, and the utilities are willing to disclose partial energy consumption data.

The calibrated urban building energy model would be most needed by municipalities, urban planners, utilities and engineering consultancies who might show keen interest to perform energy policy assessment, scenario analysis. It also has the potential to perform large-scale building performance mapping and labelling to prioritize building retrofit targets and to accelerate building stock renovation and energy transition.

Notes

1. EPISCOPE: https://episcope.eu/welcome/.
2. FME: https://www.safe.com/.
3. KNMI: https://projects.knmi.nl/klimatologie/uurgegevens/selectie.cgi.
4. EnergyPlus Weather (EPW): https://energyplus.net/weather/sources.
5. Kadaster: https://www.kadaster.nl/bag.
6. National Georegister: http://www.nationaalgeoregister.nl.
7. PDOK: https://www.pdok.nl/nl/ahn3-downloads.
8. European Building Stock Observatory: https://ec.europa.eu/energy/en/data-analysis/building-stock-observatory.
9. The TABULA WebTool: http://webtool.building-typology.eu/#bm.
10. CBS: https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische%20data/gegevens-per-postcode.
11. WoOn: https://www.wooninfo.nl/.
12. Liander open data: https://www.liander.nl/over-liander/innovatie/open-data.
13. SALib: https://salib.readthedocs.io/en/latest/.
14. Open Street Map: https://www.openstreetmap.org.
15. https://www.ep-online.nl/ep-online/Default.aspx.
16. Energy label statistics https://energiecijfers.databank.nl/dashboard/Energielabels/.
17. Amsterdam CitySim UBEM: https://github.com/ckwang25/Amsterdam_CitySim_UBEM.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Jantien Stoter received funding from the European Research Council (ERC) under the European Union's Horizon2020 Research & Innovation Programme (grant agreement no. 677312 UMnD: Urban modelling in higher dimensions).

ORCID

Cheng-Kai Wang http://orcid.org/0000-0002-8411-5758
Simon Tindemans http://orcid.org/0000-0001-8369-7568
Clayton Miller http://orcid.org/0000-0002-1186-4299
Giorgio Agugiaro http://orcid.org/0000-0002-2611-4650
Jantien Stoter http://orcid.org/0000-0002-1393-7279

References

Agentschap NL – Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. 2011. *Voorbeeldwoningen 2011: Bestaande bouw*. https://www.rvo.nl/sites/default/files/bijlagen/4.%20Brochure%20Voorbeeldwoningen%202011%20bestaande%20bouw.pdf.

Agentschap NL – Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. 2013. *Referentie woningen nieuwbouw 2013*.

Agugiaro, Giorgio, Joachim Benner, Piergiorgio Cipriano, and Romain Nouvel. 2018. "The Energy Application Domain Extension for CityGML: Enhancing Interoperability for Urban Energy Simulations." *Open Geospatial Data, Software and Standards* 3 (1): 2. doi:10.1186/s40965-018-0042-y.

Alfano, F. R. d'Ambrosio, M. Dell'Isola, G. Ficco, and F. Tassini. 2012. "Experimental Analysis of Air Tightness in Mediterranean Buildings Using the Fan Pressurization Method." *Building and Environment* 53: 16–25. doi:10.1016/j.buildenv.2011.12.017.

Allegrini, Jonas, Kristina Orehounig, Georgios Mavromatidis, Florian Ruesch, Viktor Dorer, and Ralph Evins. 2015. "A Review of Modelling Approaches and Tools for the Simulation of District-scale Energy Systems." *Renewable and Sustainable Energy Reviews* 52: 1391–1404. doi:10.1016/j.rser.2015.07.123.

Biljecki, Filip, Hugo Ledoux, and Jantien Stoter. 2014. "Height References of CityGML LOD1 Buildings and Their Influence on Applications." In: Proceedings of the ISPRS 3D GeoInfo 2014 Conference, edited by M. Breunig, et al., November 2014, Dubai, UAE. doi:10.4233/uuid:09d030b5-67d3-467b-babb-5e5ec10f1b38.

Booth, Adam T., Runchi Choudhary, and David J. Spiegelhalter. 2012. "Handling Uncertainty in Housing Stock Models." *Building and Environment* 48 (1): 35–47. doi:10.1016/j.buildenv.2011.08.016.

Bramiana, C. N., A. G. Entrop, and J. I. M. Halman. 2016. "Relationships Between Building Characteristics and Airtightness of Dutch Dwellings." *Energy Procedia* 96: 580–591. doi:10.1016/j.egypro.2016.09.103.

Cerezo, Carlos, Julia Sokol, Saud AlKhaled, Christoph Reinhart, Adil Al-Mumin, and Ali Hajiah. 2017. "Comparison of Four Building Archetype Characterization Methods in Urban Building Energy Modeling (UBEM): A Residential Case Study in Kuwait City." *Energy and Buildings* 154: 321–334. doi:10.1016/j.enbuild.2017.08.029.

Cerezo, Carlos, Julia Sokol, Christoph Reinhart, and Adil Al-Mumin. 2015. "Three Methods for Characterizing Building Archetypes in Urban Energy Simulation. A Case Study in Kuwait City." Proceedings of BS2015: 14th Conference of International Building Performance Simulation Association, Hyderabad, India, Dec 7–9, 2015.

Coakley, Daniel, Paul Raftery, and Marcus Keane. 2014. "A Review of Methods to Match Building Energy Simulation Models to Measured Data." *Renewable and Sustainable Energy Reviews* 37: 123–141. doi:10.1016/j.rser.2014.05.007.

Dukai, B., H. Ledoux, and J. E. Stoter. 2019. "A Multi-Height Lod1 Model of All Buildings in the Netherlands." ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. IV-4/W8, 2019 14th 3D GeoInfo Conference 2019, 24–27 September 2019, Singapore. doi:10.5194/isprs-annals-IV-4-W8-51-2019.

Energy Research Centre of the Netherlands. 2015. *National Energy Outlook 2015 Summary*. http://www.pbl.nl/sites/default/files/cms/publicaties/pbl-2015-dutch-national-energy-outlook-2015-summary-02325.pdf.

Environmental Systems Research Institute. 1998. *ESRI Shapefile Technical Description*. Vol. 16. http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf.

European Commission. 2015. "*EU Building Stock Observatory*." Accessed Match 10, 2018. https://ec.europa.eu/energy/en/eu-buildings-database.

Filippidou, Faidra. 2018. "Energy Performance Progress of the Dutch Non-Profit Housing Stock: A Longitudinal Assessment." PhD diss., Delft University of Technology.

Fonseca, Jimeno A., and Arno Schlueter. 2015. "Integrated Model for Characterization of Spatiotemporal Building Energy Consumption Patterns in Neighborhoods and City Districts." *Applied Energy* 142: 247–265. doi:10.1016/j.apenergy.2014.12.068.

Frayssinet, Loïc, Lucie Merlier, Frédéric Kuznik, Jean Luc Hubert, Maya Milliez, and Jean Jacques Roux. 2018. "Modeling the Heating and Cooling Energy Demand of Urban Buildings At City Scale." *Renewable and Sustainable Energy Reviews* 81: 2318–2327. doi:10.1016/j.rser.2017.06.040.

Ghiassi, Neda, and Ardeshir Mahdavi. 2017. "Reductive Bottom-up Urban Energy Computing Supported by Multivariate Cluster Analysis." *Energy and Buildings* 144: 372–386. doi:10.1016/j.enbuild.2017.03.004.

Gröger, Gerhard, Thomas Kolbe, Claus Nagel, and Karl-Heinz Häfele. 2012. *OGC City Geography Markup Language (CityGML) Encoding Standard*. https://portal.opengeospatial.org/files/?artifact_id = 47842.

Guerra Santin, Olivia, Laure Itard, and Henk Visscher. 2009. "The Effect of Occupancy and Building Characteristics on Energy Use for Space and Water Heating in Dutch Residential Stock." *Energy and Buildings* 41 (11): 1223–1232. doi:10.1016/j.enbuild.2009.07.002.

Heo, Yeonsook. 2011. "Bayesian Calibration of Building Energy Models for Energy Retrofit Decision-Making under Uncertainty." PhD diss., Georgia Institute of Technology. http://hdl.handle.net/1853/42878.

Herman, Jon, and Will Usher. 2017. "SALib: An Open-source Python Library for Sensitivity Analysis." *The Journal of Open Source Software* 2 (9): 97. doi:10.21105/joss.00097.

Howard, B., L. Parshall, J. Thompson, S. Hammer, J. Dickinson, and V. Modi. 2012. "Spatial Distribution of Urban Building Energy Consumption by End Use." *Energy and Buildings* 45: 141–151. doi:10.1016/j.enbuild.2011.10.061.

Hsu, David. 2015. "Identifying Key Variables and Interactions in Statistical Models Ofbuilding Energy Consumption Using Regularization." *Energy* 83: 144–155. doi:10.1016/j.energy.2015.02.008.

Intergovernmental Panel on Climate Change. 2007. *Climate Change 2007 Synthesis Report*.

International Energy Agency. 2013. *Transition to Sustainable Buildings – Strategies and Opportunities to 2050*. http://www.iea.org/publications/freepublications/publication/Building2013_free.pdf.

International Energy Agency. 2014. *World Energy Outlook 2014*. https://www.iea.org/publications/freepublications/publication/WEO2014.pdf.

Kavgic, M., A. Mavrogianni, D. Mumovic, A. Summerfield, Z. Stevanovic, and M. Djurovic-Petrovic. 2010. "A Review of Bottom-up Building Stock Models for Energy Consumption in the Residential Sector." *Building and Environment* 45 (7): 1683–1697. doi:10.1016/j.buildenv.2010.01.021.

Keirstead, James, Mark Jennings, and Aruna Sivakumar. 2012. "A Review of Urban Energy System Models: Approaches, Challenges and Opportunities." *Renewable and Sustainable Energy Reviews* 16 (6): 3847–3866. doi:10.1016/j.rser.2012.02.047.

Kennedy, Marc C., and Anthony O'Hagan. 2001. "Bayesian Calibration of Computer Models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63 (3): 425–464. http://doi.wiley.com/10.1111/1467-9868.00294.

Kutzner, Tatjana, and Thomas H. Kolbe. 2016. "Extending Semantic 3D City Models by Supply and Disposal Networks for Analysing the Urban Supply Situation." In *Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V.*, 13.

Ledoux, Hugo, and Martijn Meijers. 2011. "Topologically Consistent 3D City Models Obtained by Extrusion." *International Journal of Geographical Information Science* 25 (4): 557–574. doi:10.1080/13658811003623277.

Leidelmeijer, K., and P. van Grieken. 2005. *Wonen en energie – Stook en ventilatiegedrag van huishoudens*. http://publicaties.minienm.nl/download-bijlage/22305/1w477.pdf.

Loga, Tobias, Nikolaus Diefenbach, and Britta Stein. 2012. *Typology Approach for Building Stock Energy Assessment. Main Results of the TABULA project – Final Project Report*. http://www.building-typology.eu/downloads/public/docs/report/TABULA_FinalReport_AppendixVolume.pdf.

Machairas, Vasileios, Aris Tsangrassoulis, and Kleo Axarli. 2014. "Algorithms for Optimization of Building Design: A Review." *Renewable and Sustainable Energy Reviews* 31: 101–112. https://www.sciencedirect.com/science/article/abs/pii/S1364032113007855.

Mastrucci, Alessio, Olivier Baume, Francesca Stazi, and Ulrich Leopold. 2014. "Estimating Energy Savings for the Residential Building Stock of An Entire City: A GIS-based Statistical Downscaling Approach Applied to Rotterdam." *Energy and Buildings* 75: 358–367. doi:10.1016/j.enbuild.2014.02.032.

Monteiro, Claudia Sousa, Carlos Costa, André Pina, Maribel Y. Santos, and Paulo Ferrão. 2018. "An Urban Building Database (UBD) Supporting a Smart City Information System." *Energy and Buildings* 158: 244–260. doi:10.1016/j.enbuild.2017.10.009.

Nouvel, Romain, Alessio Mastrucci, Ulrich Leopold, Olivier Baume, Volker Coors, and Ursula Eicker. 2015. "Combining GIS-based Statistical and Engineering Urban Heat Consumption Models: Towards a New Framework for Multi-scale Policy Support." *Energy and Buildings* 107: 204–212. doi:10.1016/j.enbuild.2015.08.021.

Nouvel, Romain, Maryam Zirak, Volker Coors, and Ursula Eicker. 2017. "The Influence of Data Quality on Urban Heating Demand Modeling Using 3D City Models." *Computers, Environment and Urban Systems* 64: 68–80. doi:10.1016/j.compenvurbsys.2016.12.005.

Perez, Diane. 2014. "A Framework to Model and Simulate the Disaggregated Energy Flows Supplying Buildings in Urban Areas." PhD diss., École polytechnique fédérale de Lausanne. http://infoscience.epfl.ch/record/197073.

Petersdorff, Carsten, Thomas Boermans, Jochen Harnisch, Ole Stobbe, Simone Ullrich, and Sina Wartmann. 2005. *Cost-Effective Climate Protection in the EU Building Stock*. https://www.eurima.org/uploads/ModuleXtender/Publications/45/ecofysIII_report_EN.pdf.

Reinhart, Christoph F., and Carlos Cerezo Davila. 2016. "Urban Building Energy Modeling – A Review of a Nascent Field." *Building and Environment* 97: 196–202. doi:10.1016/j.buildenv.2015.12.001.

Rijksdienst voor Ondernemend Nederland. 2017. "*Energielabel Woningen*." Accessed October 5, 2018. https://www.rvo.nl/onderwerpen/duurzaam-ondernemen/gebouwen/wetten-en-regels-gebouwen/bestaande-bouw/energielabel-woningen.

Robinson, D., F. Haldi, J. Kämpf, P. Leroux, D. Perez, A. Rasheed, and U. Wilke. 2009. "CITYSIM: Comprehensive Micro-Simulation of Resource Flows for Sustainable Urban Planning." Building Simulation 2009: Eleventh International IBPSA Conference, At Glasgow, Scotland, Volume: 1083–1090.

Saltelli, Andrea, Marco Ratto, Terry Andres, Francesca Camplongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola. 2008. *Global Sensitivity Analysis: The Primer*. http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470059974.html.

Sokol, Julia, Carlos Cerezo Davila, and Christoph F. Reinhart. 2017. "Validation of a Bayesian-based Method for Defining Residential Archetypes in Urban Building Energy Models." *Energy and Buildings* 134: 11–24. doi:10.1016/j.enbuild.2016.10.050.

Swan, Lukas G., and V. Ismet Ugursal. 2009. "Modeling of End-use Energy Consumption in the Residential Sector: A Review of Modeling Techniques." *Renewable and Sustainable Energy Reviews* 13 (8): 1819–1835. doi:10.1016/j.rser.2008.09.033.

Torabi Moghadam, Sara, Jacopo Toniolo, Guglielmina Mutani, and Patrizia Lombardi. 2018. "A GIS-statistical Approach for Assessing Built Environment Energy Use At Urban Scale." *Sustainable Cities and Society* 37: (October 2017): 70–84. doi:10.1016/j.scs.2017.10.002.

United Nations. 2014. *World Urbanization Prospects: The 2014 Revision, Highlights*. https://esa.un.org/unpd/wup/publications/files/wup2014-highlights.pdf.

Wei, Tian. 2013. "A Review of Sensitivity Analysis Methods in Building Energy Analysis." *Renewable and Sustainable Energy Reviews* 20: 411–419. doi:10.1016/j.rser.2012.12.014.